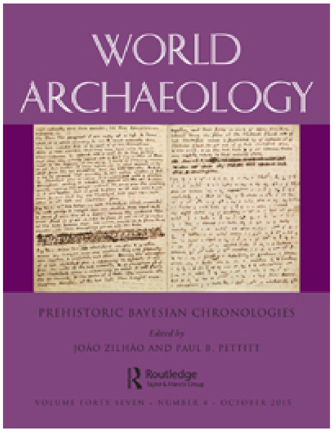

World Archaeology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rwar20

Quality in Bayesian chronological models in archaeology

Alex Bayliss[a]
[a] Historic England
Published online: 02 Sep 2015.

Quality in Bayesian chronological models in archaeology

Alex Bayliss

Abstract

Bayesian chronological modelling is fast becoming the method of choice for the interpretation of radiocarbon dates in archaeological and palaeoenvironmental studies around the world. Although software enabling the routine application of the method has been available for more than twenty years, more than half of published models have appeared in the past five years. Unfortunately, the pace of development in statistical methodology has not been matched by the increased care in sample selection and reporting that is required for robust modelling. Barely half the applications considered in this article provide the information necessary to assess the models presented critically. This article discusses what information is required to allow the quality of Bayesian chronological models to be assessed, and provides check-lists for authors, editors and referees, in the hope of improving current practice.

Keywords

Bayesian statistics; chronological modelling; sample selection; radiocarbon dating; quality assurance.

Introduction

Bayesian chronological modelling is coming of age. It is now a generation since the possibility of combining calibrated radiocarbon dates with other forms of archaeological information using Bayesian statistics was first mooted (Naylor and Smith 1988), and twenty-one years since the first software that allowed the routine application of the method was issued (Bronk Ramsey 1994, 1995).

The methodology has now been standard practice within English Heritage for more than twenty years, and within the professional archaeology sector in England for more than a decade (Bayliss and Bronk Ramsey 2004; Bayliss 2009). Forged from this body of practice,

Routledge Taylor & Francis Group *World Archaeology* Vol. 47(4): 677–700 *Prehistoric Bayesian Chronologies*
 ISSN 0043-8243 print/1470-1375 online
http://dx.doi.org/10.1080/00438243.2015.1067640

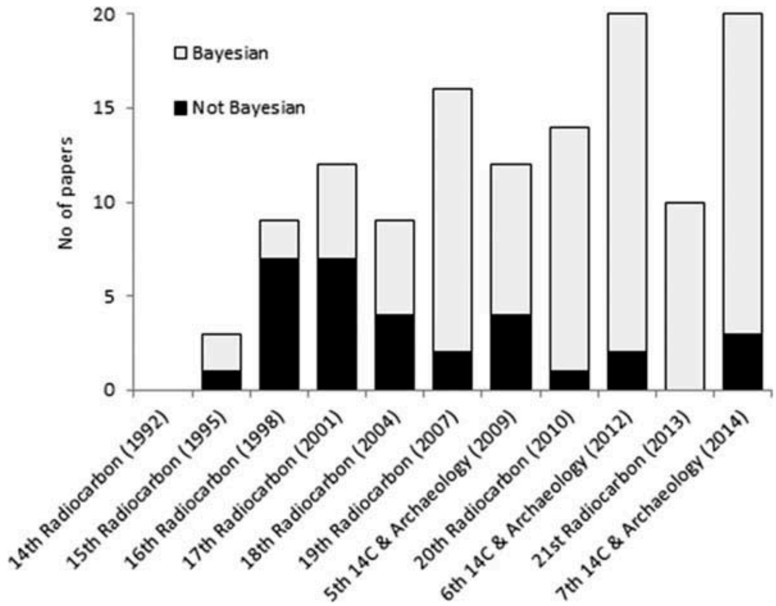

Figure 1 Number of papers including statistical approaches for the interpretation of radiocarbon dates published in the journal *Radiocarbon* arising from the 'Radiocarbon' and '^{14}C in Archaeology' conferences held since 1990.

a step-by-step process that allows the construction of rigorous chronological models on a routine basis has emerged (Bayliss 2009), along with a corpus of archaeological specialists experienced in constructing robust chronologies. This early and routine adoption of chronological modelling in England means that there is now a sufficient number of well-selected samples and well-dated sites for the first generation of generational narratives for periods of English prehistory to be constructed (Whittle, Healy and Bayliss 2011; Hamilton et al., this volume).

Internationally, Bayesian chronological modelling has been adopted more slowly, but it is now beginning to be the method of choice for the interpretation of radiocarbon dates, at least in the specialist literature. This can be illustrated, for example, by statistical methods used for the interpretation of radiocarbon dates in papers published in the journal *Radiocarbon* arising from the regular series of 'Radiocarbon' and '^{14}C in Archaeology' conferences (Fig. 1). A total of ninety-four papers employ Bayesian statistical models, of which fifty-eight (62 per cent) have been published in the past five years. In contrast, a total of thirty-one papers employ other statistical methods.[1] Of these only six (19 per cent) have appeared in the past five years. There have been two principal casualties from the rise of Bayesian statistics: the chi-squared approach to wiggle-matching (Pearson 1986), which has been replaced by a Bayesian approach that fully quantifies the error on the match (Christen and Litton 1995), and the use of summed probability distributions of calibrated radiocarbon dates (Aitchison, Ottaway and Al-Ruzaiza 1991) as the limitations of this approach have been recognized (Bayliss et al. 2007; Contreras and Meadows 2014).

In an attempt to gain a broader understanding of the use of Bayesian chronological modelling in archaeology around the world, I considered all papers published in the journals *World Archaeology, Antiquity, Radiocarbon* and the *Journal of Archaeological Science* between

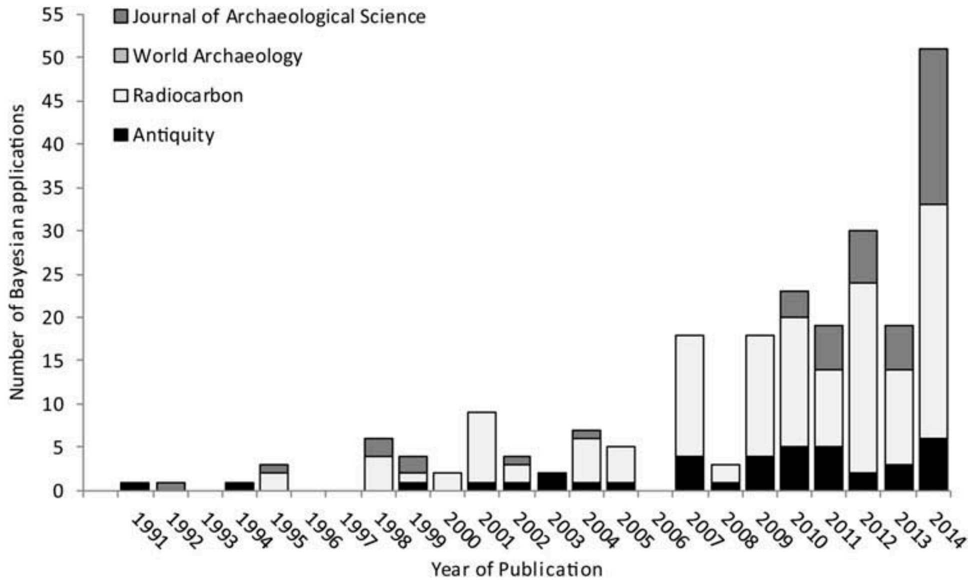

Figure 2 Number of papers including Bayesian chronological models published in the journals *Journal of Archaeological Science*, *World Archaeology*, *Radiocarbon* and *Antiquity* since 1990.

1990 and 2014. These journals are all global and multi-period in scope and so should give a representative sample of the types and number of applications being undertaken. A total of 226 papers contain Bayesian chronological models, none in *World Archaeology* (probably reflecting the thematic nature of this journal), thirty-nine in *Antiquity*, 142 in *Radiocarbon* and forty-five in the *Journal of Archaeological Science* (Fig. 2). Again, well over half of the applications, 142 (63 per cent), have been published in the last five years. Models for the chronology of individual archaeological sites predominate (60 per cent; Fig. 3), although models for the chronology of archaeological typologies (mainly pottery typologies) are also common (15 per cent), as are applications considering past environments (10 per cent).[2] These applications span the globe (Fig. 4), although the majority (88 per cent) consider Old World archaeology.

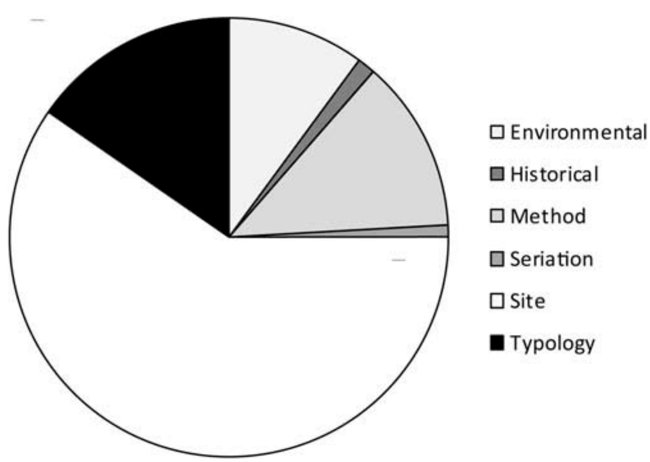

Figure 3 Types of Bayesian chronological models published in the journals *Journal of Archaeological Science, World Archaeology*, *Radiocarbon* and *Antiquity* since 1990.

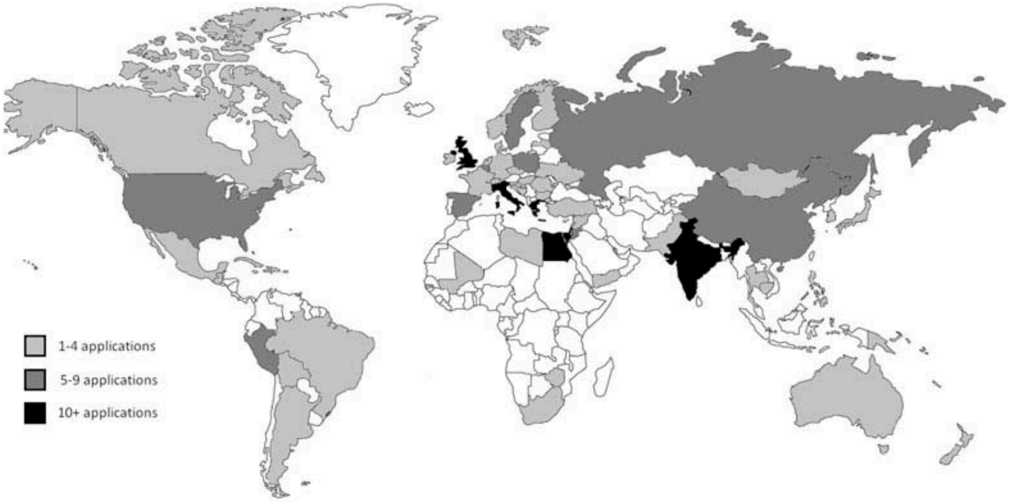

Figure 4 Numbers of sites, landscapes and archaeological typologies in different countries subject to Bayesian chronological modelling in papers published in *Journal of Archaeological Science, World Archaeology*, *Radiocarbon* and *Antiquity* since 1990.

The nature of Bayesian chronological modelling

Alison Wylie (2002, 162–3) has suggested that 'scientific arguments are more like cables than chains'. In this view individual lines of argument that are insufficient on their own can make a cumulatively persuasive case when woven together, although the strands that make up a cable of comparative, evaluative argument may conflict with one another and thus may require dynamic judgements and revisions.

In the construction of archaeological chronologies, Bayesian statistics provide a formal and explicit methodology for weaving together different strands of evidence to form the cable. Calibrated radiocarbon dates, or other date estimates on the calendar scale, are combined with archaeological prior information of various kinds to produce a combined chronology that should be more reliable than its individual components. The resultant cable should be both stronger (more robust) and tighter (more precise).

Chronological models are thus interpretative constructions. They will be revised, not only as calibration data and statistical methods improve, but also as archaeological understanding develops and as new questions are posed and more scientific dates obtained. Consequently published applications must not only explain how and why the models presented were constructed, but also provide sufficient information to allow the reader to understand the strengths and weaknesses of those models, and to allow them to be criticized and reconstructed by future researchers.

For this to be possible, it is essential that all elements of a Bayesian model are explicitly defined and discussed. It is not sufficient simply to present what has been done: the basis for each of the myriad of choices that are made during the modelling process must be outlined so that the reader can evaluate them. These choices relate to the available scientific dates and how they are included in the model, the selection of appropriate prior information and the modelling approaches employed to combine these elements and assess the reliability of resultant model. In this article, I concentrate on the first of these choices.

Defining the model

In the majority of applications considered in my review (90 per cent) the models are adequately defined. In eighteen cases (8 per cent) models are defined using formal mathematical notation. The principal focus of these studies is the introduction of new statistical methods, with applications providing exemplars of the methods discussed.

Model definition in all other cases has been undertaken using the protocols provided by the various software packages that have been used to construct the models. Specialist programs for wiggle-matching[3] simply require that the relative positions of sampled rings within the tree-ring series are specified. Specialist software for age-depth modelling[4] requires both graphical representation of the sequences and textual description of the program settings employed for explicit model definition.

Most applications, however, have been undertaken using the more flexible software packages, BCal (Buck, Christen and James 1999), Datelab (Nicholls and Jones 2002) and OxCal (Bronk Ramsey 1995, 1998, 2001, 2008, 2009a, 2009b; Bronk Ramsey, van der Plicht and Weninger 2001; Bronk Ramsey et al. 2010; Bronk Ramsey and Lee 2013). Models constructed using BCal are described using the algebraic notation for model definition provided by that software (e.g. Acabado 2009), and the few models constructed using Datelab were defined by textual description of the model structure (e.g. Jacomb et al. 2014). Around three-quarters of the chronological models in my sample, however, have been defined using the Chronological Query Language provided by OxCal (Bronk Ramsey 1998, 2009a). Most usually this is done using the CQL keywords and brackets shown on the left-hand side of the diagrams produced by the program (e.g. Hey, Bayliss and Boyle 1999 for an example of the CQL coding used by OxCal v.1–v.3 and Valzolgher et al. 2012 for an example of the CQL2 coding using by OxCal v.4), although models are also defined by publishing the program code (e.g. Marsh 2012) or by describing this in the text (e.g. Guo et al. 2001). Many papers make a use of a mixture of these techniques, or redefine components of models that have been fully described elsewhere (e.g. Parker Pearson et al. 2007, fig. 6, which redefines a component of a model for Stonehenge which is otherwise described by Bronk Ramsey and Bayliss (2000, figs 5.7–5.8).

A few papers simply fail to define explicitly the model(s) discussed, although a number of others attempt to do so but provide insufficient information.[5] Where models are defined using protocols specific to the software package employed, it is essential that the software (including the version used) is specified.

Reporting the data (scientific criteria)

Details of the radiocarbon measurements, the methods used to produce them, and the samples which were analysed are essential information for the construction and subsequent evaluation of any Bayesian chronological model.

Unfortunately, protocols for calculating and reporting radiocarbon determinations were agreed by the international radiocarbon community over thirty-five years ago (Stuiver and Polach 1977) and have not been revised substantively since then (Stuiver 1980, 1983; Mook 1986). Andrew Millard (2014) has proposed sorely needed updated conventions, although these have yet to be ratified and cover only the scientific information that must be reported for each measurement. Millard's proposals do not currently include the archaeological information that needs to be reported for each radiocarbon sample if it is to be incorporated in Bayesian chronological models either at the time of commission or in subsequent research.

For all radiocarbon measurements, the fractionation-corrected radiocarbon age and associated error term should be published along with the unique laboratory identifier.[6] For modern samples, the fractionation-corrected result and associated error term should be reported.[7]

Details of associated measurements should also be provided. The measurements provided vary both by the sampled material and by the dating laboratory. Most common by far are $\delta^{13}C$ values, which are used for fractionation correction in the calculation of radiocarbon ages. Unfortunately, at present there is no consensus among radiocarbon laboratories about how $\delta^{13}C$ is measured and about which values are reported to users. Consequently, it is necessary to record both how the $\delta^{13}C$ value that has been used to calculate the conventional radiocarbon age has been measured and how the $\delta^{13}C$ value that has been reported for publication has been measured.[8] Occasionally, a sample may be too small to allow a $\delta^{13}C$ value to be measured by conventional mass spectrometry and an assumed value appropriate to the relevant material might be used for age calculation. This was much more common in the past when AMS machines did not measure $\delta^{13}C$ values on-line, and many radiocarbon dating facilities did not have access to conventional mass spectrometry.

Following $\delta^{13}C$ values, the most common associated measurements are C:N ratios, $\delta^{15}N$ values, and percentage yield for bone samples and percent carbon values for charred plant material. These associated measurements provide quality assurance for the reported radiocarbon measurements, but can also provide vital information for their interpretation (e.g. by indicating the possibility of marine or freshwater reservoir effects).

Details of the facility/facilities which produced the results and how samples were pre-treated, prepared for measurement and dated should also be provided. This is particularly important in cases where it is not apparent from the laboratory code whether a sample was dated by AMS or by liquid scintillation spectrometry (e.g. Wk-, Beta-). In cases where different chemical fractions of a sample can be selected for dating, it is important that it is clear exactly which fraction has been combusted and dated (e.g. for bulk sediments whether the alkali-soluble/acid-insoluble, 'humic acid', or acid and alkali insoluble, 'humin', fraction has been dated). References to published papers are ideal as these will be traceable by the next generation of researchers once methods have advanced and current dating facilities have ceased operation.

The final type of scientific information that is required in assessing the inputs into a Bayesian chronological model relates to the calibration of the radiocarbon measurements. Although, in Bayesian modelling, calibration becomes part of the modelling process, it is still an essential step. Details of the calibration curve used for model calculation, along with any reservoir offset used (including the error on the offset), and the resolution at which the model has been calculated must therefore all be published as part of model definition.

A selection of examples of technical information published for samples of a range of materials from a range of archaeological sites, all taken from the sample of papers considered in this review, is provided in Table 1 (see Table 2 for descriptions of these samples). A check-list for publication of the scientific information on radiocarbon dates needed for the construction (and de-construction) of Bayesian chronological models is provided in Table 3.

Generally, the definition of scientific dating in the sample of papers considered here is of a high standard (although it is possible that this is not an accurate reflection of current practice, since so many of these papers were published in specialist scientific journals *Radiocarbon* and *Journal of Archaeological Science*). Nearly three-quarters of applications provide a table of the radiocarbon ages included in the models (with the associated errors and laboratory codes), another 17 per cent refer to data that are published elsewhere and in only 9 per cent of cases are relevant data not provided at all.

As part of the modelling process all dates have been calibrated, and the required technical details regarding the calibration data and any reservoir corrections used have been provided. Practice varies as to which calendar date estimates are cited in publication. Around half of papers publish simple calibrated date ranges for each radiocarbon measurement, although it is often not clear which method has been used for calibration. It is also not clear what purpose these calibrated date ranges serve. They are not the most accurate indication of the date of particular samples, as this is the Highest Posterior Density interval from the relevant Bayesian model. But this is rarely given, even when it may be meaningful, for example when considering the dates of particular graves (e.g. OxA-13251, Sample 1 in Tables 1 and 2, which is the date of Grave 112 at Varna). Most frequently only the Highest Posterior Density intervals of key parameters from a model are provided, for example the *start* and *end* parameters for the Varna cemetery (Higham et al. 2007, figs 3–4).

Calibration and indeed the posterior density estimates from today's models are fundamentally disposable. It is unhelpful to the reader if these are not presented clearly, but ultimately any model has a shelf-life that is limited by the next iteration of the internationally agreed calibration curve. I have myself, for example, recently re-calculated the model for the Varna cemetery presented by Higham et al. (2007, fig. 3), with updated calibration data (Tasić et al. 2015, fig. 13). But the basic scientific data are essential in this process of re-modelling and, although the radiocarbon ages are usually reported and laboratory methods are often referenced, too frequently associated measurements such as $\delta^{13}C$ values are missing and it is a rare application that reports the fraction of a bulk sediment sample that was dated.

Defining the data (archaeological criteria)

Publishing the radiocarbon results and associated scientific data in full is only the first part of the information required to assess the quality and treatment of data in a Bayesian chronological model. Details of the samples that were processed by the laboratory are paramount.

First, we must assess the material that was dated. Was it a short-life sample which obtained its carbon from a well-understood carbon reservoir? Basically, the species and maturity of every sample should be obtained before it is sent for dating, and that information must be published. Unfortunately, however, almost a quarter of all the published applications considered in this article fail to provide information that allows the reader to assess whether a sample consisted of short-lived material. In the remaining three-quarters of applications, short-lived samples are dominated by samples of human and animal bone, and in a significant proportion of papers

Table 1 Reporting radiocarbon and stable isotope measurements

Laboratory number	Sample details (see Table 3)	Radiocarbon age (BP)	$\delta^{13}C$ (AMS) (‰)	$\delta^{13}C$ (IRMS) (‰)	$\delta^{15}N$ (IRMS) (‰)	C:N (atomic)	%C	% collagen (by weight)
OxA-13251	Sample 1	5702±32		−18.6	10.5	3.1	39.9	1.2
KIA-40095	Sample 2	4816±26	−18.0±0.2					
OxA-22380	Sample 3	2516±26		−19.4	8.7	3.2	44.5	0.7
SNU05-040	Sample 4A	890±40		−17.2				
SNU05-537	Sample 4B	670±90		−19.5				
Weighted mean: Sample 4, T′ = 4.9; ν = 1; T′(5%) = 3.8, 855 ± 37 BP								
SNU05-536	Sample 5	770±200		−25.7		0.5		
SNU05-538	Sample 6	790±40		−25.8		0.5		
KIA-45802	Sample 7	4575±35		−22.8±0.2	6.4±0.2			
OxA-18010	Sample 8	28,650±200		−18.4±0.2	3.4±0.3	3.2	42.6	4.2
K-4255	Sample 9	3380±61		−23.8				
AA-78973	Sample 10	119±38		−25.2				
Wk-8206	Sample 11	733±16		−26.4±0.2				
KIA-45764	Sample 12	2745±30						
AA-52359	Sample 13	1065±40		−26.0				
LU-5315	Sample 14	8810±120						
OxA-20253	Sample 15	34,980±220		−1.8				

Note Replicate measurements have been tested for consistency and combined as described by Ward and Wilson (1978).

Table 2 Reporting sample details

Sample	*Sample details*		
Sample 1	Human bone from extended inhumation in Grave 112 from the cemetery at Varna, Bulgaria, accompanied by a copper needle, a flint blade, a miniature trapezoidal polished limestone adze, a heavily destroyed deer antler tool, a necklace composed of three metamorphosed ultrabasite and eleven *Spondylus* beads, five clay vessels and small fragments of a ceramic sieve (Higham et al. 2007, fig. 2: 1–7).	Human bone from richly furnished Grave 112 at Varna	Varna Grave 112
Sample 2	001, disarticulated human ?clavicle from F8000, use of Dolmen II at Flintbek LA 3, Germany (Mischka 2011).	Human bone from F8000, Dolmen II at Flintbek LA 3	Human bone from Flintbek LA 3
Sample 3	Articulating animal bone, *Sus scrofa*, deposited as a mortuary offering in Burial 56 EPV at Ban Chiang, Thailand (Higham, Higham and Kijngam 2011).	Pig bone from Burial 56 EPV at Ban Chiang	Animal bone from Ban Chiang Burial 56 EPV
Sample 4A	Tarsal bone from an articulating skeleton of a *c*. 40-year-old woman in grave 1 (2004), accompanied by a headless horse skeleton with a gilded saddle and a gold ring from Tavan Tolgoi, Mongolia (Youn et al. 2007).	Human bone from female skeleton in grave 1 (2004) from Tavan Tolgoi	Bone from grave 1 (2004) from Tavan Tolgoi
Sample 4B	Two fragments of human rib bone from an articulating skeleton of a *c*. 40-year-old woman in grave 1 (2004); replicate of samples 21A.	Replicate of sample 21A	Bone from grave 1 (2004) from Tavan Tolgoi
Sample 5	Finely woven textile (outer garment) from a *c*. 40-year-old female skeleton in grave 1 (2004), accompanied by a headless horse skeleton with a gilded saddle and a gold ring from Tavan Tolgoi, Mongolia (Youn et al. 2007).	Textile from outer garment of woman buried in grave 1 (2004) from Tavan Tolgoi	Textile from grave 1 (2004) from Tavan Tolgoi
Sample 6	Fragments of textile adhered to the inner surface of the hip bone of an articulating skeleton of a *c*. 40-year-old female skeleton in grave 1 (2004), accompanied by a headless horse skeleton with a gilded saddle and a gold ring from Tavan Tolgoi, Mongolia (Youn et al. 2007).	Textile from inner garment of woman buried in grave 1 (2004) from Tavan Tolgoi	Textile from grave 1 (2004) from Tavan Tolgoi
Sample 7	Cattle (*Bos* sp.) left proximal tibia with refitting unfused epiphysis from layer 10, trench 1, midden sequence from Riņņukalns, Latvia (Bērziņš et al. 2014).	Cattle tibia from midden layer 10 at Riņņukalns	Animal bone from layer 10 at Riņņukalns
Sample 8	Reindeer (*Rangifer tarandus*) cut radius from context M10 from the Champ de Fouilles, Maisières-Canal, Belgium (Jacobi et al. 2010)	Reindeer radius from context M10 at Maisières-Canal	Animal bone from Maisières-Canal

(*continued*)

Table 2 (*Continued*)

Sample	*Sample details*		
Sample 9	Charred twig (*Tamarix* sp.) from House 3, Delta 1, from volcanic destruction layer on Thera, Greece (Manning et al. 2002).	*Tamarix* sp. charcoal from destruction layer of House 3, Thera.	Charcoal from Thera destruction layer
Sample 10	Fragment of charcoal, *Pinus kesiya* Royle ex Gordon, from hard earth fill and wall foundation (layer II), Mamag excavation unit, of the Bocos terrace system, Banaue, Ifugao, the Philippines (Acabado 2009).	Charcoal, *Pinus kesiya* Royle ex Gordon, from layer II of the Bocos terrace system	Charcoal from Bocos terrace system
Sample 11	Rings 21–30 from carbonized *Phyllocladus* spp. log, 0.15m in diameter with bark intact, from near-source pyroclastic deposits (unit Hpdc) formed early in the Kaharoa eruption episode; sampled at the Crater Road section near Mount Tarawera, New Zealand (Hogg et al. 2003).	Rings 21–30 from carbonized *Phyllocladus* spp. log buried in the Kaharoa eruption	Carbonized log buried in the Kaharoa eruption
Sample 12	SU 108, bulk sample of six charred cereal grains from a collapse layer (containing material from construction) in the East Outside, on bedrock and abutting walls SU 1047, SU 1044, SU 1032, and SU 1037 from Cornia Nou, Menorca, Spain (Anglada et al. 2014).	SU 108, charred grain from collapse layer in East Outside from Cornia Nou	Grain from collapse layer from Cornia Nou
Sample 13	S.2, carbonized food crust on exterior of London shellyware sherd from C.4715 at 75 High Street, Perth, Scotland (Hall et al. 2007).	Carbonized food crust on London shellyware sherd from 75 High Street, Perth	Carbonized residue from 75 High Street, Perth
Sample 14	Bulk peat (*Carex* sp.) from a depth of 540–60cm from a sediment core at Nizhne-Psinovskoe, Russia; ‘humin’ fraction dated (Dolukhanov et al. 2007).	*Carex* peat from 540–560cm from Nizhne-Psinovskoe	Peat from Nizhne-Psinovskoe
Sample 15	Marine shell (Fra 8), *Cyclopes* sp., from stratum Q in FAS 222 in Franchthi Cave, Greece (Douka et al. 2011).	Marine shell, *Cyclopes* sp., from stratum Q in FAS 222 in Franchthi Cave	Shell from stratum Q in Franchthi Cave

Table 3 Check-list for authors, editors and referees for reporting radiocarbon and associated measurements

Information	*Example (following Hogg et al. 2003)*
Laboratory number	Wk-8206
Radiocarbon age/modern value and error	733±16 BP
Calculation details	Stuiver and Polach 1977, corrected for fractionation using reported value
References to laboratory methods: pre-treatment, synthesis and measurement	Acid-base-acid pre-treatment with the NaOH step carried out in a N atmosphere; liquid scintillation spectrometry of benzene optimized for high-precision measurement (Higham and Hogg 1997; McCormac et al. 1998a, 1998b; Hogg 1992).
Associated measurements including how reported $\delta^{13}C$ values were measured (and errors if reported)	−26.4±0.2‰ (IRMS)
Calibration data used, including any reservoir offsets (with errors)	Calibrated using the Southern Hemisphere atmospheric radiocarbon calibration curve (Hogg et al. 2002)

details are provided of samples that were probably of long-lived material. Charcoal samples are particularly poorly served, with the majority being unidentified.

The good, the bad and the ugly of sample description are illustrated in Table 2. If we consider, for example, Sample 9, the description in the left-hand column tells us that the sample was of short-lived material: it was a twig. The description of the same sample in the middle column does not tell us this. It tells us that the sample was *Tamarisk* sp., but nothing about its maturity. Most plants of this genus are shrubs, and so we might infer that this sample was likely to have incorporated an age-at-death offset of no more than a few decades. But this is an inference, not an observation. The description of this sample in the right-hand column is even worse. The charcoal could have been a tamarisk twig and so a short-life sample, but it could equally have been from a juniper trunk and incorporated an age-offset of several centuries. In each column in Table 2, Sample 9 is accurately described: it was a twig, it was tamarisk, and it was charcoal. But our assessment of the quality of the Bayesian chronological modelling presented in this paper (Manning et al. 2002) and the utility of this result in future research is fundamentally affected by the reporting of the material dated.

The vast majority of samples dated are clearly from the terrestrial radiocarbon reservoir, although a few applications report measurements on samples of marine shell or samples with more complex or variable carbon reservoirs, such as lime mortar or freshwater shell. In recent years, there has been increasing awareness of the potential for dietary offsets in the dating of human bone (e.g. Bērziņš et al. 2014; Bronk Ramsey et al. 2014), although there are also case studies where pairs of contemporary human bone and fully terrestrial material indicate no dietary offsets (e.g. Higham et al. 2007; Youn et al. 2007). Carbon and nitrogen stable isotopic values are increasingly reported for samples of human bone in an attempt to identify cases where dietary offset may occur. Generally, the carbon reservoir of the dated samples is explicitly considered in published applications, since this information is required to select appropriate calibration options.

The next characteristic of the dated material that may affect the interpretation of a date within a chronological model is whether it was a bulk sample or a single entity (Ashmore 1999). The importance of this information can be illustrated using simple statistics. Consider, for example, a deposit where one in ten of the recovered charred seeds is residual (reworked from an earlier context). If we date a single seed from this deposit, the radiocarbon date will have a 90 per cent chance of actually dating to the time when the context was formed and a 10 per cent chance of being earlier. If we obtain two radiocarbon dates from this deposit, each from a single seed, then there will be a 99 per cent chance that at least one of the two dates will relate to the time when the deposit was formed. If, however, we bulk those two seeds together and obtain one radiocarbon date, then there will be a 19 per cent chance that at least one of those seeds is residual and so the radiocarbon date is earlier than the time when the deposit was formed. The greater the number of items that are bulked together, the lower the probability that the sample will contain only freshly deposited material. If ten seeds were to be bulked together for dating from this deposit, then there would be a chance of less than 1 in 3 that the resultant radiocarbon date would accurately date its formation. Obviously, the scale of the offset will depend on the actual proportion of residual material in a sample and its date in relation to the time when the deposit was formed. But, all else being equal, a bulk sample will have a higher chance of containing reworked material than a single-entity sample. In chronological modelling this might be reflected, for example, in the prior outlier probabilities given to each type of sample in an outlier analysis.

The importance of single-entity sampling does not appear to be widely appreciated. Nearly 40 per cent of applications do not provide the information necessary to determine whether a sample was a single entity and in many of the remainder samples are reported as single-entity or not incidentally, either because they are samples of human bone from inhumations or because they are radiometric measurements on materials which would have to be bulked to obtain a large enough sample for conventional dating (e.g. charred grain). It is rare that a sample of charred material dated by AMS is explicitly recorded as being a single entity (although it should be noted that some species of seed have to be bulked to obtain sufficient material even for AMS dating). It should be noted that both ‘grain’ and ‘seed’ are collective as well as singular nouns in English. Consider, for example, Sample 12 in Table 2. The description in the left-hand column clear states that this was a ‘bulk sample of six charred cereal grains’, but the descriptions in the central and right-hand columns are ambivalent about whether a single charred cereal grain or a bag of charred grain was combusted and dated.

The final characteristic of the dated material that is of importance in utilizing radiocarbon dates in Bayesian chronological modelling is whether a sample derives from a different organism from those that produced the other samples. This determines whether a group of measurements are statistically independent. This is particularly important when estimating the duration of a period of activity. If the same organism has been dated multiple times (and this is not known and the measurements combined before inclusion in the model), then the model may be biased towards a shorter estimate of duration. Often the independence of samples is unknown but improbable. For example, if we date two different grains of barley from a charred store of grain, then it is possible that the barley grains came from the same ear of barley, but this is improbable since the store of grain will have derived from a whole field of barley. If, however, we are sampling from a limited number of human corpses buried in a collective burial monument, then the chance of sampling the same individual more than once is much higher.

For example, Sample 2 in Table 2 is recorded as 'a disarticulated human ?clavicle'. Knowing this, if we wished to obtain further dates from this tomb that we knew were from distinct individuals, we could take samples from other clavicles (it would be better if we knew whether a left or right clavicle had been sampled initially).

Having assessed how the characteristics of the dated material affect our assessment of the quality of a Bayesian model, we now require information on the association of that dated material with the archaeological problem that the model addresses (Waterbolk 1971). This relationship, between the *dated event* (e.g. the shedding of an antler) and the *target event* (e.g. the digging of a Neolithic ditch), is never known but is inferred on the basis of archaeological evidence. The basis of this inference, and its security, must be specifically considered for every sample and the modelling approach selected accordingly.

The most secure association is when the dated material comes from an object that is of intrinsic interest. An example is provided by Sample 13 in Table 2. Here the sample is a carbonized food crust adhering to a diagnostic sherd of London shellyware from Perth, Scotland. If the objective of the dating programme is to obtain a chronology for the use of London shellyware, then it would not matter if this sample was unstratified. Similarly, in wiggle-matching it is the sequence of tree-rings within the sample itself that is important, although the archaeological importance of the precise dating provided by the wiggle-match may derive from the association of the dated timber with an archaeological event (such as the eruption of Kaharoa dated by Sample 11 in Table 2).

Cases where it is the date of the samples themselves that is of interest, however, are comparatively rare. It is usually the context of the sample that is of interest: the date of the ditch, or the site, or the associated material culture. This is all the more important if we have a stratigraphic sequence of deposits that we wish to use as prior information in a Bayesian chronological model. Stratigraphy, of course, provides evidence about the relative sequence of contexts. Radiocarbon dating does not date contexts, it dates samples. So, we can constrain the calibrated radiocarbon dates using the stratigraphic sequence of units only if the samples were freshly deposited in the context from which they were recovered. The crucial archaeological interpretation is to establish whether a potential sample is likely to have been residual (or, less frequently, intrusive) in the context from which it was recovered. This may be inferred with varying degrees of confidence.

Sometimes we may have physical evidence that a sample is contemporary with the activity that we wish to date. This might take the form of cut-marks on animal bones in a prehistoric hunting camp (e.g. Sample 8 in Table 2) or articulated bones in a grave (e.g. Samples 1,4A and 4B in Table 2). Sometimes the physical characteristics of samples may suggest that they are unlikely to be residual or intrusive. Articulated or articulating bone groups (e.g. Sample 7 in Table 2) fall into this category, as do residues on groups of refitting pottery sherds or shells of bi-valves where both halves are present. Waterlogged material in isolated deposits, such as the base of a well, would probably not have been preserved in a waterlogged condition if it was not in its original context. Less certainly, the fragility of certain types of samples might be taken to imply that reworking is unlikely.

In other cases, we might infer a functional relationship between the material that is dated and the context from which it derived. Such might be a calcined bone in cremation or charcoal from a hearth. Structural material might survive, such as charred timbers from buildings that were destroyed by fire or waterlogged wattle panels. The tools putatively associated with construction

Table 4 Check-list for authors, editors and referees for reporting radiocarbon samples

Information	*Example (following Bērziņš et al. 2014, KIA-45802)*
Reservoir	Northern Hemisphere atmospheric (Reimer et al. 2013)
Age-at-death	Animal bone
Single-entity	Cattle left proximal tibia with refitting unfused epiphysis
Distinct individual	Cattle left tibia
Association (physical or functional)	Cattle left proximal tibia with refitting unfused epiphysis from layer 10, trench 1 midden sequence, Riņņukalns, Latvia (Bērziņš et al. 2014, fig. 6)

events, such as antler picks from the base of prehistoric ditches, may also be found. In other cases the association may be more problematic. Was a bone in a grave-fill placed there deliberately as a grave good? If so, was it an heirloom? Was a coherent, friable deposit rich in charred plant remains deposited as a single event? How did an 'occupation' deposit on a house floor accumulate? What was the origin of the datable material recovered? Was it trampled in from outside? Or overspill from the central hearth? Or from the make-up that constituted the fabric of the floor itself?

There are no hard-and-fast rules or perfect samples. All samples contain an element of risk, and it is ultimately a matter of archaeological judgement as to which material is submitted for dating. But this thread is critical in the weave of a Bayesian model and must be open to evaluation. For this to be possible, the detail which forms the basis of each archaeological judgement must be presented. For example, a fragment of short-lived charcoal selected from 20g of charcoal (>2mm) from a black rake-out deposit (20L) adjacent to Hearth 1 might convincingly be argued to be fuel relating to the use of the hearth. A fragment of short-lived charcoal selected from ten fragments (>2mm) from the reddened base of Hearth 1 might be a less convincing functional association (did the charcoal derive from the clay make-up of the hearth?).

The provision of the archaeological information needed to assess the quality of Bayesian models is of a lower standard than the provision of the scientific information considered above. Although over three-quarters of applications record the context from which dated samples came, often this is simply a record of the unit number. Only slightly more than half of papers discuss the association of the dated samples with the archaeological problem that is being modelled, and in some cases even this is a generic discussion rather than one concerning the particular samples that were dated.

A check-list for publication of the archaeological information on radiocarbon dates needed for the construction (and de-construction) of Bayesian chronological models is provided in Table 4.

Data definition and model (de)construction

Overall, of the 226 papers considered in this review, only 102 (45 per cent) provide all the information required to assess the quality of the chronological model(s) presented (as summarized in Tables 3 and 4).

I would like to conclude by considering why this matters. How does the detail of the data affect the choices made in model construction and in our assessment of the quality of the

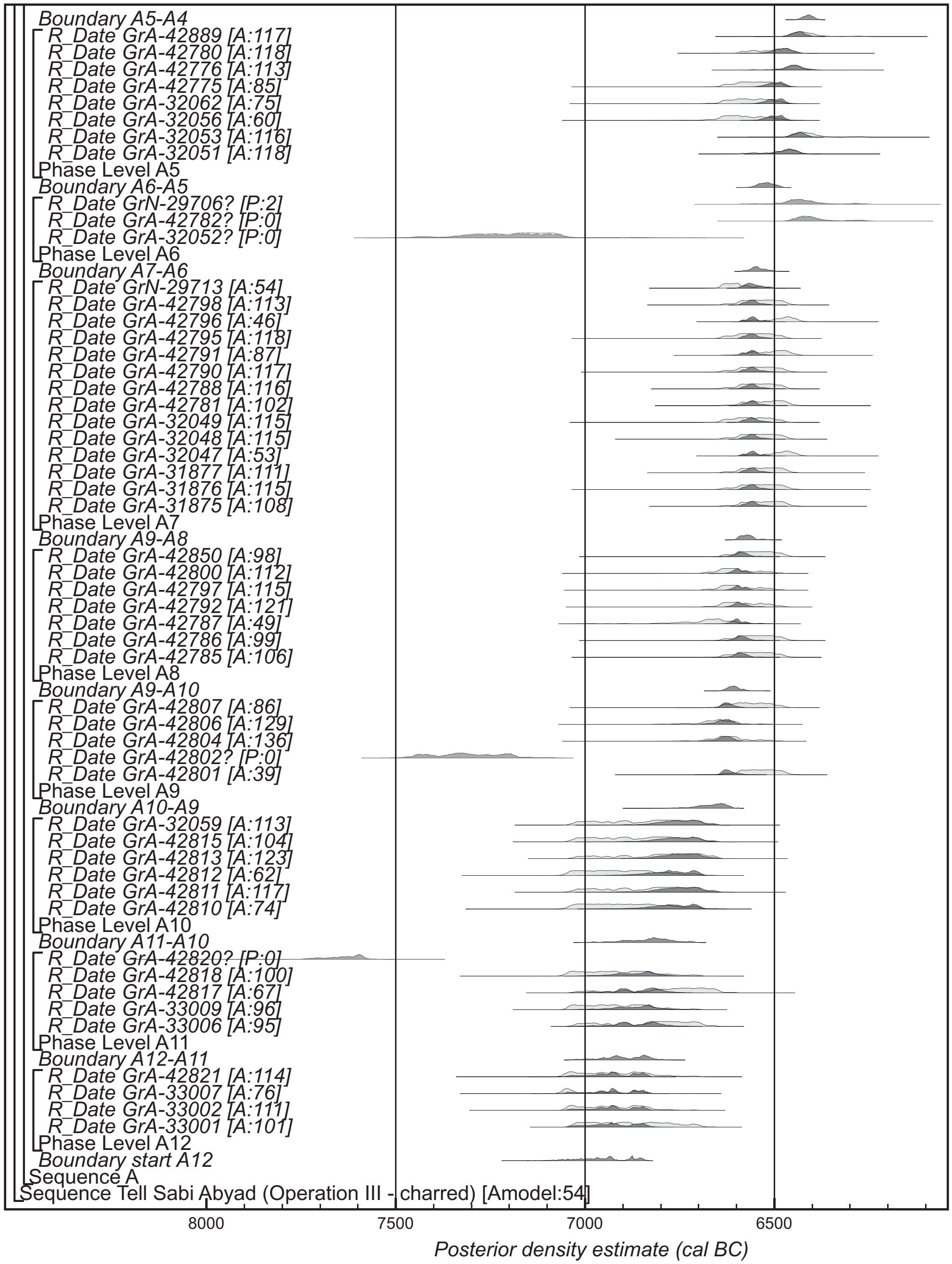

Figure 5 Probability distributions of dates on human burials from Tell Sabi Abyad, Operation III. Each distribution represents the relative probability that an event occurred at a particular time. For each of the dates two distributions have been plotted, one in outline which is the result produced by the scientific evidence alone and a solid one which is based on the chronological model used. The other distributions correspond to aspects of the model. For example, the distribution '*start A12*'' is the estimated date when Level A12 on the site started. Dates followed by a question mark have been calibrated (Stuiver and Reimer 1993), but not included in the chronological model for reasons explained in the text. Those in grey have been identified as outlier by van der Plicht et al. (2011) or by re-analysis. The model is defined by the keywords and brackets on the left-hand side of Figs 5–7 (basal component).

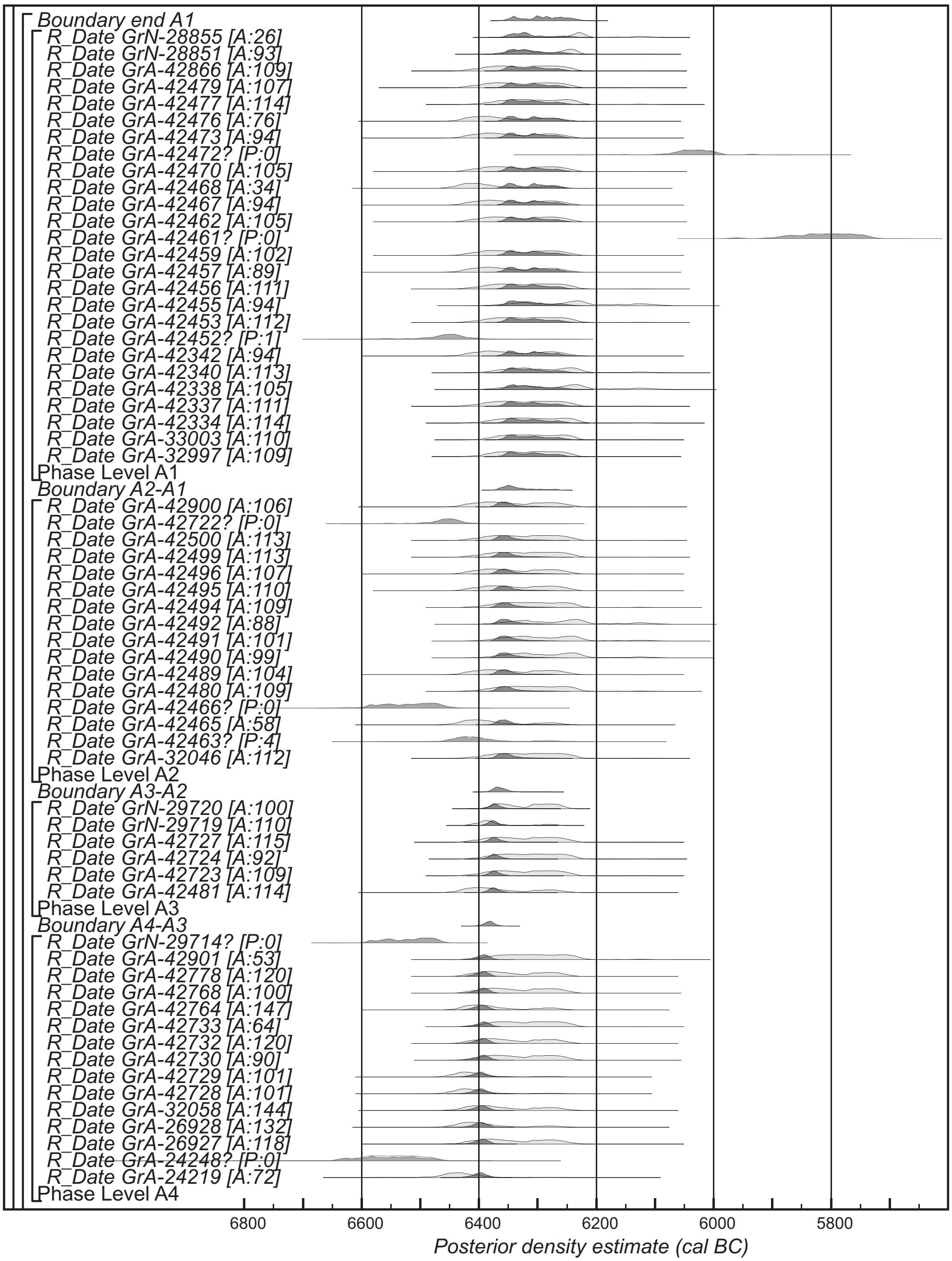

Figure 6 Probability distributions of dates on charred material from Tell Sabi Abyad, Operation III. The format is identical to that of Figure 5. The model is defined by the keywords and brackets on the left-hand side of Figures 5–7 (middle component).

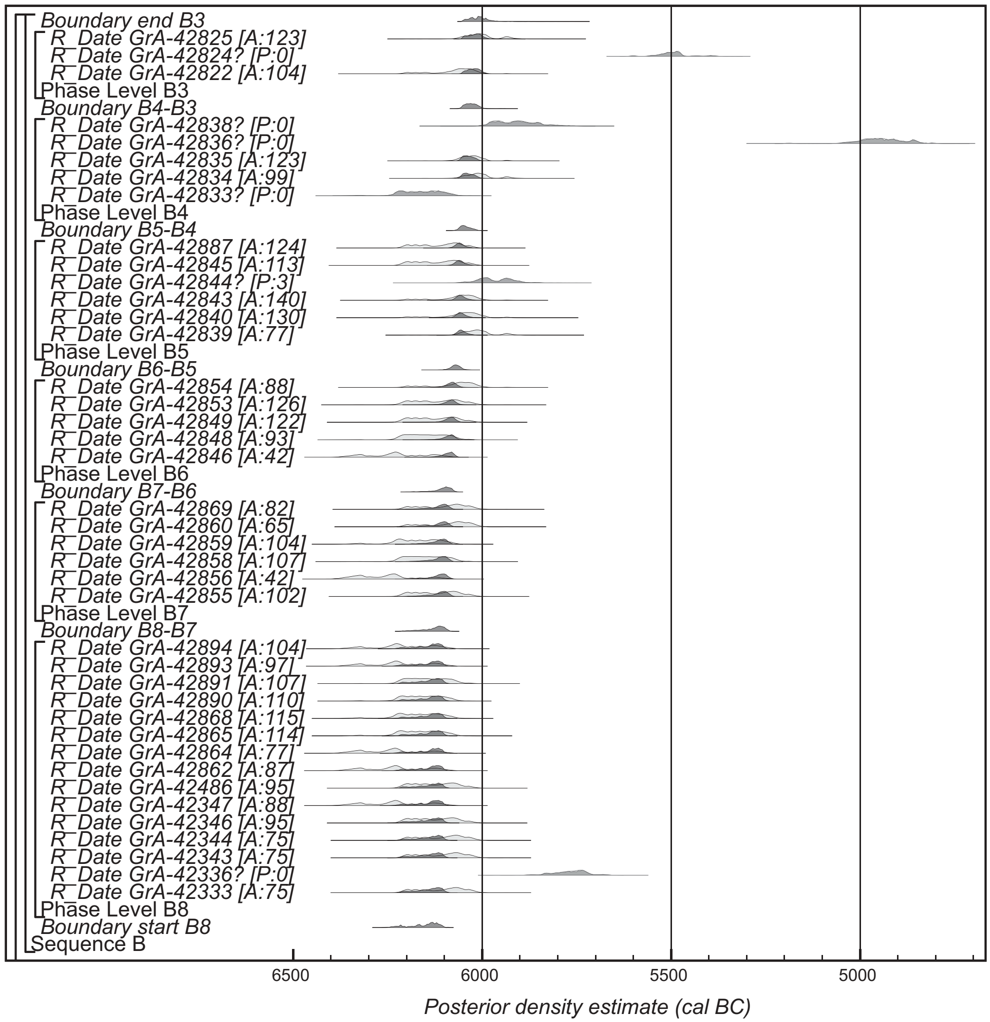

Figure 7 Probability distributions of dates on charred material from Tell Sabi Abyad, Operation III. The format is identical to that of Figure 5. The model is defined by the keywords and brackets on the left-hand side of Figures 5–7 (upper component).

chronological models presented? To illustrate this discussion I consider two interim models for the chronology of Tell Sabi Abyad, Syria.[9]

The first model is based on a suite of 163 radiocarbon dates on charred plant material from the lower part of the structural sequence in Operation III, of which 145 dates were included in a Bayesian chronological model (van der Plicht et al. 2011). The structure of this model is described clearly in the text and I have been able to reconstruct it using the data provided in the Appendix.[10] Unfortunately, the data are in such poor agreement with this model that it will not calculate. I have therefore manually identified a number of additional outliers to produce a workable model (Figs 5–7).[11]

It is here that the limitations of this paper appear. Detailed sample descriptions are not provided in the data appendix, rather there is a generic statement about sample character:

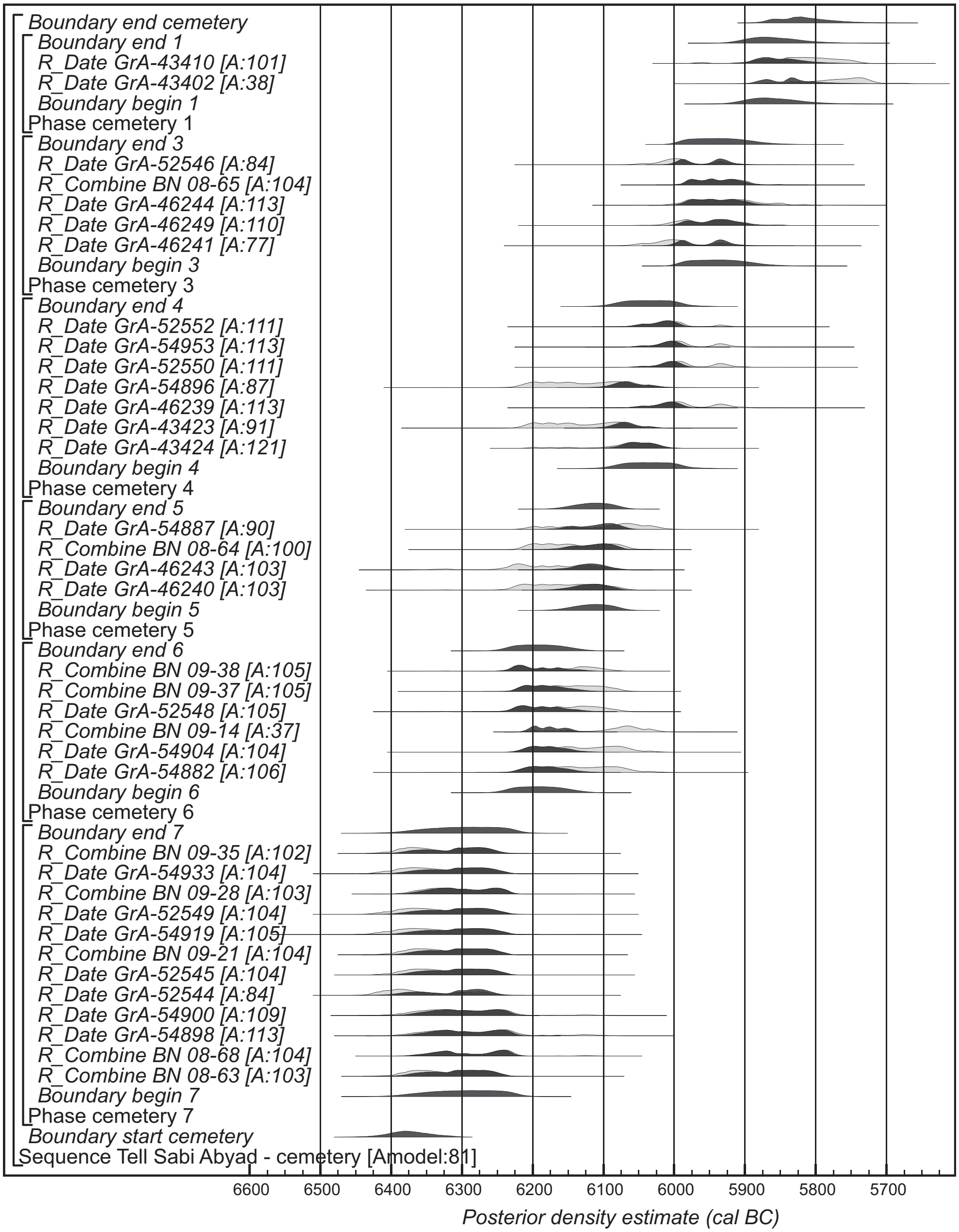

Figure 8 Probability distributions of dates on human burials from Tell Sabi Abyad, Operation III (burials). The format is identical to that of Figure 5. The model is defined by the keywords and brackets on the left-hand side of this figure.

> Practically all samples designated as charcoal in the date list represent in fact unidentified seeds/grains, shrubs, and twigs, i.e. short-lived sample material. Possible ‘old wood effects’ are not an issue here. The samples come from a closed context: collected from bins, fills, ovens, hearths, and rooms. (van der Plicht et al. 2011, 232)

Given the clear outliers in these data, both these statements cannot always be true. GrA-42820 in Phase A11, for example, is clearly either a charcoal sample from a substantial timber or residual.

Further details of the character of each sample and the context from which it came would allow us to decide exactly how to incorporate each date into the model (for example, grain samples might be fully incorporated, samples of unidentified charcoal incorporated as *termini post quos* or using the charcoal outlier function of OxCal; Dee and Bronk Ramsey 2014). More fundamentally, maybe it is the overall site phasing into 'Levels' that requires a critical re-examination. Perhaps the data would be more compatible with a sequence based purely on stratigraphic superposition. As it stands, we have enough information to judge that the chronological model presented in this paper must be viewed with some scepticism, but not enough to be able to suggest an alternative reading.

The second model from Tell Sabi Abyad is from the late Neolithic cemetery, also from Operation III. Plug, van der Plicht and Akkermans (2014) present a model in which forty-six measurements on articulated skeletons have good agreement with the stratigraphic sequence (Fig. 8, Amodel: 81; after Plug, van der Plicht and Akkermans 2014, fig. 5). Strenuous attempts have been made to ensure the accuracy of measurements on bone, on a site with very poor collagen preservation. These include a suite of associated measurements (%C, %N, C:N ratios), and no fewer than ten statistically consistent pairs of replicate measurements. In this case, not only do we have sufficient information to be confident that the presented model is plausible, but we also have a candid account of the limitations of the reading of the stratigraphy currently incorporated in the model (Plug, van der Plicht and Akkermans 2014, 546).

Conclusion

Accurate and detailed reporting of the samples selected for radiocarbon dating and the measurements produced is fundamental to assessing the quality of the Bayesian chronological models reported and, probably more fundamentally, in reusing those data to produce alternative readings of the archaeological problems considered. Current practice is generally flawed, and renewed attention should be paid to the reporting of both the scientific and archaeological aspects of the raw data upon which all our models are founded.

Acknowledgements

Many thanks to Johannes van der Plicht for fruitlessly searching for the original OxCal code for the Tell Sabi Abyad model published in 2011.

Disclosure statement

No potential conflict of interest was reported by the author.

Alex Bayliss
Historic England
alex.bayliss@historicengland.org.uk

Notes

1 Not including simple calibration of radiocarbon dates and chi-squared tests to determine the statistical consistency of replicate groups of measurements.
2 Environmental applications, especially age-depth models of deposits at particular locations, are probably more common than appears from this sample as many such applications are published in specialist environmental journals.
3 Bwigg (Christen 2003) and RH3.2w (Imamura 2007).
4 BPeat (Blaauw and Christen 2005) and BACON (Blaauw and Christen 2011) are included in my sample of applications.
5 It should be noted that both the brackets and the keywords in an OxCal diagram are required for explicit definition of the model.
6 A list of internationally agreed laboratory codes is available at http://radiocarbon.org/Info/labcodes.html. Care should be taken to ensure that it is these codes (i.e. SUERC-, NZA-), and not internal laboratory tracking numbers (i.e. GU-, R-), that are published with the measurements.
7 In the absence of agreed protocols, the methods used for the calculation of the modern result should be referenced, as current practice varies (e.g. Reimer, Brown and Reimer 2004; Mook and van der Plicht 1999).
8 Some laboratories use $\delta^{13}C$ values measured on the AMS to calculate ages, and report those values (e.g. KIA-); some laboratories use $\delta^{13}C$ values measured by conventional isotopic ratio mass spectrometry (IRMS) to calculated ages, and report those values (e.g. SUERC-); some laboratories use $\delta^{13}C$ values measured on the AMS to calculate ages, but report a second $\delta^{13}C$ value on the same sample measured by conventional mass spectrometry (e.g. OxA-, GrA-); some laboratories use $\delta^{13}C$ values measured on the AMS to calculate ages, but do not routinely report them (e.g. Poz-). In some laboratories, practice has varied through time (e.g. UBA-).
9 Both models are interim statements of the chronology of parts of the site, since it is very likely that the phasing sequences incorporated in the models will be modified during the final post-excavation analysis that is currently under way. The fact that I have been able to use these papers as exemplars is a tribute to the quality of their reporting (despite the caveats discussed in the text).
10 The only exception to this is in Phase B8 where a second outlier is excluded from the original analysis, which I have not been able to identify from the information presented.
11 The models described in this section have been constructed in OxCal v4.2.4, and calculated at a resolution of 5 using IntCal13 (Reimer et al. 2013).

References

Acabado, S. 2009. "A Bayesian Approach to Dating Agricultural Terraces: A Case from the Philippines." *Antiquity* 83: 801–14. doi:10.1017/S0003598X00099002.

Aitchison, T., B. Ottaway, and A. S. Al-Ruzaiza. 1991. "Summarizing a Group of ^{14}C Dates on the Historical Time Scale: With a Worked Example from the Late Neolithic of Bavaria." *Antiquity* 65: 108–16.

Anglada, M., A. Ferrer, L. Plantalamor, D. Ramis, M. Van Strydonck, and G. De Mulder. 2014. "Chronological Framework for the Early Talayotic Period in Menorca: The Settlement of Cornia Nou." *Radiocarbon* 56: 411–24. doi:10.2458/56.16962.

Ashmore, P. 1999. "Radiocarbon Dating: Avoiding Errors by Avoiding Mixed Samples." *Antiquity* 73: 124–30.

Bayliss, A. 2009. "Rolling Out Revolution: Using Radiocarbon Dating in Archaeology." *Radiocarbon* 51: 123–47.

Bayliss, A., and C. Bronk Ramsey. 2004. "Pragmatic Bayesians: A Decade of Integrating Radiocarbon Dates into Chronological Models." In *Tools for Constructing Chronologies: Tools for Crossing Disciplinary Boundaries*, edited by C. E. Buck and A. R. Millard, 25–41. Berlin: Springer Verlag.

Bayliss, A., C. Bronk Ramsey, J. van der Plicht, and A. Whittle. 2007. "Bradshaw and Bayes: Towards a Timetable for the Neolithic." *Cambridge Archaeological Journal* 17: 1–28.

Bērziņš, V., U. Brinker, C. Klein, H. Lübke, J. Meadows, M. Rudzīte, U. Schmölcke, H. Stümpel, and I. Zagorska. 2014. "New Research at Riņņukalns, a Neolithic Freshwater Shell Midden in Northern Latvia." *Antiquity* 88: 715–32. doi:10.1017/S0003598X0005064X.

Blaauw, M., and J. A. Christen. 2005. "Radiocarbon Peat Chronologies and Environmental Change." *Applied Statistics* 54: 805–16.

Blaauw, M., and J. A. Christen. 2011. "Flexible Paleoclimate Age-Depth Models Using an Autoregressive Gamma Process." *Bayesian Analysis* 6: 457–74. doi:10.1214/ba/1339616472.

Bronk Ramsey, C. 1995. "Calibration and Analysis of Stratigraphy: The Oxcal Program." *Radiocarbon* 36: 425–30.

Bronk Ramsey, C. 1998. "Probability and Dating." *Radiocarbon* 40: 461–74.

Bronk Ramsey, C. 2001. "Development of the Radiocarbon Calibration Program Oxcal." *Radiocarbon* 43: 355–63.

Bronk Ramsey, C. 2008. "Deposition Models for Chronological Records." *Quaternary Science Reviews* 27: 42–60. doi:10.1016/j.quascirev.2007.01.019.

Bronk Ramsey, C. 2009a. "Bayesian Analysis of Radiocarbon Dates." *Radiocarbon* 51: 37–60.

Bronk Ramsey, C. 2009b. "Dealing with Outliers and Offsets in Radiocarbon Dating." *Radiocarbon* 51: 1023–45.

Bronk Ramsey, C., and A. Bayliss. 2000. "Dating Stonehenge." In *CAA96: Computer Applications and Quantitative Methods in Archaeology*, edited by K. Lockyer, T. J. T. Sly, and V. Mihailescu-Birliba, BAR Int Ser, 845, 29–39. Oxford: Archaeopress.

Bronk Ramsey, C., M. Dee, T. Nakagawa, and R. Staff. 2010. "Developments in the Calibration and Modelling of Radiocarbon Dates." *Radiocarbon* 52: 953–61.

Bronk Ramsey, C., and S. Lee. 2013. "Recent and Planned Developments of the Program Oxcal." *Radiocarbon* 55: 720–30. doi:10.2458/azu_js_rc.55.16215.

Bronk Ramsey, C., R. Schulting, O. I. Goriunova, V. I. Bazaliiskii, and A. J. Weber. 2014. "Analyzing Radiocarbon Reservoir Offsets through Stable Nitrogen Isotopes and Bayesian Modeling: A Case Study

Using Paired Human and Faunal Remains from the Cis-Baikal Region, Siberia." *Radiocarbon* 56: 789–99. doi:10.2458/56.17160.

Bronk Ramsey, C., J. van der Plicht, and B. Weninger. 2001. "'Wiggle Matching' Radiocarbon Dates." *Radiocarbon* 43: 381–89.

Buck, C. E., J. A. Christen, and G. N. James. 1999. "Bcal: An Online Bayesian Radiocarbon Calibration Tool." *Internet Archaeology* 7. doi:10.11141/issn.1363-5387.

Christen, J. A. 2003. "Bwigg: An Internet Facility for Bayesian Radiocarbon Wiggle-Matching." *Internet Archaeol* 13. doi:10.11141/1a.13.2.

Christen, J. A., and C. D. Litton. 1995. "A Bayesian Approach to Wiggle-Matching." *Journal of Archaeological Science* 22: 719–25. doi:10.1016/0305-4403(95)90002-0.

Contreras, D. A., and J. Meadows. 2014. "Summed Radiocarbon Calibrations as A Population Proxy: A Critical Evaluation Using A Realistic Simulation Approach." *Journal of Archaeological Science* 52: 591–608. doi:10.1016/j.jas.2014.05.030.

Dee, M., and C. Bronk Ramsey. 2014. "High-Precision Bayesian Modeling of Samples Susceptible to Inbuilt Age." *Radiocarbon* 56: 83–94. doi:10.2458/56.16685.

Dolukhanov, P. M., A. M. Shukurov, K. A. Arslanov, D. A. Subetto, G. I. Zaitseva, E. N. Djinoridze, D. D. Kuznetsov, A. V. Ludikova, T. V. Sapelko, and L. A. Savelieva. 2007. "Evolution of Waterways and Early Human Settlements in the Eastern Baltic Area: Radiocarbon-Based Chronology." *Radiocarbon* 49: 527–42.

Douka, K., C. Perlès, H. Valladas, M. Vanhaeren, and R. E. M. Hedges. 2011. "Franchthi Cave Revisited: The Age of the Aurignacian in South-Eastern Europe." *Antiquity* 85: 1131–50. doi:10.1017/S0003598X00061962.

Guo, Z., K. Liu, X. Lu, H. Ma, K. Li, J. Yuan, S. Yuan, X. Wu, and L. Xu. 2001. "AMS Radiocarbon Dating of Tianma-Qucun Site in Shanxi, China." *Radiocarbon* 43: 1109–14.

Hall, D. W., G. T. Cook, M. A. Hall, G. K. P. Muir, D. Hamilton, and E. M. Scott. 2007. "The Early Medieval Origin of Perth, Scotland." *Radiocarbon* 49: 639–44.

Hamilton, et al. *This Volume*.

Hey, G., A. Bayliss, and A. Boyle. 1999. "Iron Age Inhumation Burials at Yarnton, Oxfordshire." *Antiquity* 73: 551–62.

Higham, C., T. Higham, and A. Kijngam. 2011. "Cutting a Gordian Knot: The Bronze Age of Southeast Asia: Origins, Timing and Impact." *Antiquity* 85: 583–98. doi:10.1017/S0003598X00067971.

Higham, T., J. Chapman, V. Slavchev, B. Gaydarska, N. Honch, Y. Yordanov, and B. Dimitrova. 2007. "New Perspectives on the Varna Cemetery (Bulgaria) – AMS Dates and Social Implications." *Antiquity* 81: 640–54. doi:10.1017/S0003598X00095636.

Higham, T. F. G., and A. G. Hogg. 1997. "Evidence for Late Polynesian Colonisation of New Zealand: University of Waikato Radiocarbon Measurements." *Radiocarbon* 39: 149–92.

Hogg, A. G. 1992. "Performance and Design of 0.3ml to 10ml Synthetic Silica Liquid Scintillation Vials for Low-Level ^{14}C Determination." *Radiocarbon* 34: 135–42.

Hogg, A. G., T. F. G. Higham, D. J. Lowe, J. G. Palmer, P. J. Reimer, and R. M. Newnham. 2003. "A Wiggle-Match Date for Polynesian Settlement of New Zealand." *Antiquity* 77: 116–25. doi:10.1017/S0003598X00061408.

Hogg, A. G., F. G. McCormac, T. F. G. Higham, P. J. Reimer, M. G. L. Baillie, J. G. Palmer, and M. Stuiver. 2002. "High-Precision ^{14}C Measurements of Contemporaneous Tree-Ring Dated Wood from the British Isles and New Zealand: AD 1850–950." *Radiocarbon* 44: 633–40.

Imamura, M. 2007. "On the Radiocarbon Calibration Program File, RHC3.2." *Bulletin of the National Museum of Japanese History* 137 ((in Japanese)).

Jacobi, R. M., T. F. G. Higham, P. Haesaerts, I. Jadin, and L. S. Basell. 2010. “Radiocarbon Chronology for the Early Gravettian of Northern Europe: New AMS Determinations for Maisières-Canal, Belgium.” *Antiquity* 84: 26–40. doi:10.1017/S0003598X00099749.

Jacomb, C., R. N. Holdaway, M. E. Allentoft, M. Bunce, C. L. Oskam, R. Walters, and E. Brooks. 2014. “High-Precision Dating and Ancient DNA Profiling of Moa (Aves: Dinornithiformes) Eggshell Documents a Complex Feature at Wairau Bar and Refines the Chronology of New Zealand Settlement by Polynesians.” *Journal of Archaeological Science* 50: 24–30. doi:10.1016/j.jas.2014.05.023.

Manning, S., C. Bronk Ramsey, C. Doumas, T. Marketou, G. Cadogan, and C. L. Pearson. 2002. “New Evidence for an Early Date for the Aegean Late Bronze Age and Thera Eruption.” *Antiquity* 76: 733–44. doi:10.1017/S0003598X00091183.

Marsh, E. J. 2012. “A Bayesian Re-Assessment of the Earliest Radiocarbon Dates from Tiwanaku, Bolivia.” *Radiocarbon* 54: 203–18. doi:10.2458/azu_js_rc.v54i2.15826.

McCormac, F. G., A. G. Hogg, T. F. G. Higham, M. G. L. Baillie, J. G. Palmer, L. Xiong, J. R. Pilcher, D. Brown, and S. T. Hoper. 1998a. “Variations of Radiocarbon in Tree-Rings: Southern Hemisphere Offset Preliminary Results.” *Radiocarbon* 40: 1153–59.

McCormac, F. G., A. G. Hogg, T. F. G. Higham, J. Lynch-Stieglitz, W. S. Broecker, M. G. L. Baillie, J. G. Palmer, *et al*. 1998b. “Temporal Variation in the Interhemispheric ^{14}C Offset.” *Geophysical Research Letters* 25: 1321–24. doi:10.1029/98GL01065.

Millard, A. 2014. “Conventions for Reporting Radiocarbon Determinations.” *Radiocarbon* 56: 555–59. doi:10.2458/56.17455.

Mischka, D. 2011. “The Neolithic Burial Sequence at Flintbek La 3, North Germany, and Its Cart Tracks: A Precise Chronology.” *Antiquity* 85: 742–58. doi:10.1017/S0003598X00068289.

Mook, W., and J. van der Plicht. 1999. “Reporting ^{14}C Activities and Concentrations.” *Radiocarbon* 41: 227−39.

Mook, W. G. 1986. “Business Meeting: Recommendations/Resolutions Adopted by the Twelfth International Radiocarbon Conference.” *Radiocarbon* 28: 799.

Naylor, J. C., and A. F. M. Smith. 1988. “An Archaeological Inference Problem.” *Jounal of American Statistical Association* 83: 588–95.

Nicholls, G., and M. Jones. 2002. “New Radiocarbon Calibration Software.” *Radiocarbon* 44: 663–74.

Parker Pearson, M., R. Cleal, P. Marshall, S. Needham, J. Pollard, C. Richards, C. Ruggles, *et al*. 2007. “The Age of Stonehenge.” *Antiquity* 81: 617–39. doi:10.1017/S0003598X00095624.

Pearson, G. W. 1986. “Precise Calendrical Dating of Known Growth-Period Samples, Using a ‘Curve-Fitting’ Technique.” *Radiocarbon* 28: 292–99.

Plug, H., J. van der Plicht, and P. M. M. G. Akkermans. 2014. “Tell Sabi Abyad, Syria: Dating of Neolithic Cemeteries.” *Radiocarbon* 56: 543–54. doi:10.2458/56.17446.

Ramsey, B. 1994. “Analysis of Chronological Information and Radiocarbon Calibration: The Program OxCal.” *Archaeological Computing Newsletter* 41: 11–16.

Reimer, P. J., E. Bard, A. Bayliss, J. W. Beck, P. Blackwell, C. Bronk Ramsey, C. E. Buck, *et al*. 2013. “Intcal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years Cal BP.” *Radiocarbon* 55: 1869–87. doi:10.2458/azu_js_rc.55.16947.

Reimer, P. J., T. A. Brown, and R. W. Reimer. 2004. “Discussion: Reporting and Calibration of Post-Bomb ^{14}C Data.” *Radiocarbon* 46: 1299−1304.

Stuiver, M. 1980. “Workshop on ^{14}C Data Reporting.” *Radiocarbon* 22: 964–66.

Stuiver, M. 1983. “Business Meeting: International Agreements and the Use of the New Oxalic Acid Standard.” *Radiocarbon* 25: 793–95.

Stuiver, M., and H. A. Polach. 1977. “Reporting of ^{14}C Data.” *Radiocarbon* 19: 355–63.

Stuiver, M., and P. J. Reimer. 1993. “Extended ^{14}c Database and Revised CALIB 3.0 ^{14}c Age Calibration Program.” *Radiocarbon* 35: 215–230.

Tasić, N., M. Marić, K. Penezić, D. Filipović, K. Borojević, N. Russell, P. Reimer, et al. 2015. “The End of the Affair: Formal Chronological Modelling for the Top of the Neolithic Tell of Vinča-Belo Brdo.” *Antiquity*.

Valzolgher, E., J. Meadows, P. Salzani, and L. Salzani. 2012. “Radiocarbon Dating of the Early Bronze Age Cemetery at Arano, Verona, Northern Italy.” *Radiocarbon* 54: 483–503.

van der Plicht, J., P. M. M. G. Akkermans, O. Nieuwenhuyse, A. Kaneda, and A. Russell. 2011. “Tell Sabi Abyad, Syria: Radiocarbon Chronology, Cultural Change, and the 8.2ka Event.” *Radiocarbon* 53: 229–43.

Ward, G. K., and S. R. Wilson. 1978. “Procedures for Comparing and Combining Radiocarbon Age Determinations – A Critique.” *Archaeometry* 20: 19–31.

Waterbolk, H. T. 1971. “Working with Radiocarbon Dates.” *Proceedings of the Prehistoric Society* 37: 15–33. doi:10.1017/S0079497X00012548.

Whittle, A., F. Healy, and A. Bayliss. 2011. *Gathering Time: Dating the Early Neolithic Enclosures of Southern Britain and Ireland*. Oxford: Oxbow.

Wylie, A. 2002. *Thinking with Things: Essays in the Philosophy of Archaeology*. Berkeley, CA: University of California Press.

Youn, M., J. C. Kim, H. K. Kim, D. Tumen, D. Navaan, and M. Erdene. 2007. “Dating the Tavan Tolgoi Site, Mongolia: Burial of the Nobility from Genghis Khan’s Era.” *Radiocarbon* 49: 685–91.

Alex Bayliss is Head of Scientific Dating at Historic England, and Professor of Archaeological Science at the University of Stirling. She has pioneered the routine use of Bayesian statistics for chronological modelling, constructing hundreds of models from thousands of radiocarbon dates.